

DKRUSE

#11

BP

OIPE

9-27-01

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/453,387A

DATE: 07/25/2001

TIME: 15:26:46

Input Set : A:\Uco-956.app

Output Set: N:\CRF3\07252001\I453387A.raw

PS

3 <110> APPLICANT: Wilkins, Thea A.  
 4 The Regents of the University of California  
 6 <120> TITLE OF INVENTION: Cotton Transcription Factors and Their Uses  
 8 <130> FILE REFERENCE: 023070-095600US  
 10 <140> CURRENT APPLICATION NUMBER: US 09/453,387A  
 11 <141> CURRENT FILING DATE: 1999-12-02  
 13 <160> NUMBER OF SEQ ID NOS: 26  
 15 <170> SOFTWARE: PatentIn Ver. 2.1  
 17 <210> SEQ ID NO: 1  
 18 <211> LENGTH: 1006  
 19 <212> TYPE: DNA  
 20 <213> ORGANISM: Gossypium hirsutum  
 22 <220> FEATURE:  
 23 <221> NAME/KEY: CDS  
 24 <222> LOCATION: (59)..(943)  
 25 <223> OTHER INFORMATION: GhMYB1  
 27 <400> SEQUENCE: 1  
 28 taacaccgtt attctttctc tattctacct gatttgattt gatggattt tgtaactg 58  
 30 atg gga cga tca cct tgt tgt gaa aag gct cat acc aac aaa ggt gcc 106  
 31 Met Gly Arg Ser Pro Cys Cys Glu Lys Ala His Thr Asn Lys Gly Ala  
 32 1 5 10 15  
 34 tgg acc aaa gag gaa gat caa cgc ctc atc aac tac atc cgt gtc cat 154  
 35 Trp Thr Lys Glu Glu Asp Gln Arg Leu Ile Asn Tyr Ile Arg Val His  
 36 20 25 30  
 38 ggt gaa ggc tgc tgg cgt tcc ctc ccc aaa gct gct ggg ctg ctt aga 202  
 39 Gly Glu Gly Cys Trp Arg Ser Leu Pro Lys Ala Ala Gly Leu Leu Arg  
 40 35 40 45  
 42 tgt ggt aag agt tgc aga tta aga tgg ata aac tac ttg agg cct gat 250  
 43 Cys Gly Lys Ser Cys Arg Leu Arg Trp Ile Asn Tyr Leu Arg Pro Asp  
 44 50 55 60  
 46 ctt aag aga gga aat ttc act gaa gaa gat gag ctt atc atc aag 298  
 47 Leu Lys Arg Gly Asn Phe Thr Glu Glu Asp Glu Leu Ile Ile Lys  
 48 65 70 75 80  
 50 ctt cac agt tta ctt gga aac aaa tgg tca ttg att gct gga aga tta 346  
 51 Leu His Ser Leu Leu Gly Asn Lys Trp Ser Leu Ile Ala Gly Arg Leu  
 52 85 90 95  
 54 cca gga aga aca gat aat gag ata aag aac tac tgg aac aca cac atc 394  
 55 Pro Gly Arg Thr Asp Asn Glu Ile Lys Asn Tyr Trp Asn Thr His Ile  
 56 100 105 110  
 58 aaa aga aag ctt ata agc aga gga att gat cca caa act cat cgt cct 442  
 59 Lys Arg Lys Leu Ile Ser Arg Gly Ile Asp Pro Gln Thr His Arg Pro  
 60 115 120 125  
 62 ctc aat caa acg gcc aat acc aac aca gtc aca gcc ccc acc gaa ttg 490  
 63 Leu Asn Gln Thr Ala Asn Thr Val Thr Ala Pro Thr Glu Leu  
 64 130 135 140  
 66 gat ttc aga aac tcg ccc aca tcc gtt tcc aaa tcc agt tcc atc aaa 538  
 67 Asp Phe Arg Asn Ser Pro Thr Ser Val Ser Lys Ser Ser Ser Ile Lys

ENTERED

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/09/453,387A

DATE: 07/25/2001  
TIME: 15:26:46

Input Set : A:\Uco-956.app  
Output Set: N:\CRF3\07252001\I453387A.raw

68	145	150	155	160	
70	aac ccg tct ctg gat ttc aat tac aat gaa ttt caa ttc aag tcc aac				586
71	Asn Pro Ser Leu Asp Phe Asn Tyr Asn Glu Phe Gln Phe Lys Ser Asn				
72	165	170	175		
74	aca gat tcc ctt gaa gaa ccc aac tgt aca gcc agc agt ggc atg act				634
75	Thr Asp Ser Leu Glu Glu Pro Asn Cys Thr Ala Ser Ser Gly Met Thr				
76	180	185	190		
78	aca gat gaa gag caa caa gaa cag ctg cac aag aag cag caa tac ggt				682
79	Thr Asp Glu Glu Gln Gln Glu Glu His Lys Lys Gln Gln Tyr Gly				
80	195	200	205		
82	ccg agc aat ggg caa gac ata aat ttg gag ctg tcg att ggg att gtt				730
83	Pro Ser Asn Gly Gln Asp Ile Asn Leu Glu Leu Ser Ile Gly Ile Val				
84	210	215	220		
86	tca gct gac tca tct cgg gta tca aat gcc aac tcg gcc gag tcg aaa				778
87	Ser Ala Asp Ser Ser Arg Val Ser Asn Ala Asn Ser Ala Glu Ser Lys				
88	225	230	235	240	
90	cca aag gta gat aac aac aat ttc cag ttt ctt gaa caa gct atg gtg				826
91	Pro Lys Val Asp Asn Asn Asn Phe Gln Phe Leu Glu Gln Ala Met Val				
92	245	250	255		
94	gct aag gcg gta tgt ttg tgt tgg caa tta ggt ttt gga aca agt gaa				874
95	Ala Lys Ala Val Cys Leu Cys Trp Gln Leu Gly Phe Gly Thr Ser Glu				
96	260	265	270		
98	att tgt agg aac tgt caa aat tca aat tca aat ggc ttc tat agt tat				922
99	Ile Cys Arg Asn Cys Gln Asn Ser Asn Ser Asn Gly Phe Tyr Ser Tyr				
100	275	280	285		
102	tgt aga ccc ttg gat tca tag ggtcatcttt ttcttcttcc tttctgtttt				973
103	Cys Arg Pro Leu Asp Ser				
<b>W--&gt;</b>	<b>104</b>	<b>290</b>	<b>295</b>		
106	taggagataa attaatgctt aattataaaa aaa				1006
109	<210> SEQ ID NO: 2				
110	<211> LENGTH: 294				
111	<212> TYPE: PRT				
112	<213> ORGANISM: Gossypium hirsutum				
114	<400> SEQUENCE: 2				
115	Met Gly Arg Ser Pro Cys Cys Glu Lys Ala His Thr Asn Lys Gly Ala				
116	1	5	10	15	
117	Trp Thr Lys Glu Glu Asp Gln Arg Leu Ile Asn Tyr Ile Arg Val His				
118	20	25	30		
119	Gly Glu Gly Cys Trp Arg Ser Leu Pro Lys Ala Ala Gly Leu Leu Arg				
120	35	40	45		
121	Cys Gly Lys Ser Cys Arg Leu Arg Trp Ile Asn Tyr Leu Arg Pro Asp				
122	50	55	60		
123	Leu Lys Arg Gly Asn Phe Thr Glu Glu Asp Glu Leu Ile Ile Lys				
124	65	70	75	80	
125	Leu His Ser Leu Leu Gly Asn Lys Trp Ser Leu Ile Ala Gly Arg Leu				
126	85	90	95		
127	Pro Gly Arg Thr Asp Asn Glu Ile Lys Asn Tyr Trp Asn Thr His Ile				
128	100	105	110		
129	Lys Arg Lys Leu Ile Ser Arg Gly Ile Asp Pro Gln Thr His Arg Pro				

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/453,387A

DATE: 07/25/2001

TIME: 15:26:46

Input Set : A:\Uco-956.app

Output Set: N:\CREF3\07252001\I453387A.raw

130	115	120	125
131	Leu Asn Gln Thr Ala Asn Thr Asn Thr Val Thr Ala Pro Thr Glu Leu		
132	130	135	140
133	Asp Phe Arg Asn Ser Pro Thr Ser Val Ser Lys Ser Ser Ser Ile Lys		
134	145	150	155
135	Asn Pro Ser Leu Asp Phe Asn Tyr Asn Glu Phe Gln Phe Lys Ser Asn		160
136	165	170	175
137	Thr Asp Ser Leu Glu Glu Pro Asn Cys Thr Ala Ser Ser Gly Met Thr		
138	180	185	190
139	Thr Asp Glu Glu Gln Gln Glu Gln Leu His Lys Lys Gln Gln Tyr Gly		
140	195	200	205
141	Pro Ser Asn Gly Gln Asp Ile Asn Leu Glu Leu Ser Ile Gly Ile Val		
142	210	215	220
143	Ser Ala Asp Ser Ser Arg Val Ser Asn Ala Asn Ser Ala Glu Ser Lys		
144	225	230	235
145	Pro Lys Val Asp Asn Asn Asn Phe Gln Phe Leu Glu Gln Ala Met Val		240
146	245	250	255
147	Ala Lys Ala Val Cys Leu Cys Trp Gln Leu Gly Phe Gly Thr Ser Glu		
148	260	265	270
149	Ile Cys Arg Asn Cys Gln Asn Ser Asn Ser Asn Gly Phe Tyr Ser Tyr		
150	275	280	285
151	Cys Arg Pro Leu Asp Ser		
152	290		
156	<210> SEQ ID NO: 3		
157	<211> LENGTH: 1151		
158	<212> TYPE: DNA		
159	<213> ORGANISM: Gossypium hirsutum		
161	<220> FEATURE:		
162	<221> NAME/KEY: CDS		
163	<222> LOCATION: (72)..(752)		
164	<223> OTHER INFORMATION: GhMYB6		
166	<400> SEQUENCE: 3		
167	cggattttct ttccccgtgt ttgggtgcac agaaaagttag agaaaagttt acttttgatt 60		
169	ttgaaactcc g atg aga aaa cct tgc tgc gat aaa caa ggc acc aac aag 110		
170	Met Arg Lys Pro Cys Cys Asp Lys Gln Gly Thr Asn Lys		
171	1 5 10		
173	gga gcc tgg tcc aag caa gaa gat caa aag ctc att gat tat ata cgt 158		
174	Gly Ala Trp Ser Lys Gln Glu Asp Gln Lys Leu Ile Asp Tyr Ile Arg		
175	15 20 25		
177	att cat ggt gaa ggc tgt tgg cgt tcc ctc ccc aaa gct gca ggt ttg 206		
178	Ile His Gly Glu Gly Cys Trp Arg Ser Leu Pro Lys Ala Ala Gly Leu		
179	30 35 40 45		
181	cac cgt tgc ggt aaa agt tgc agg ctg aga tgg ata aat tac tta aga 254		
182	His Arg Cys Gly Lys Ser Cys Arg Leu Arg Trp Ile Asn Tyr Leu Arg		
183	50 55 60		
185	cca gat atc aaa cgt ggt aac ttt gct caa gac gaa gag gac tta att 302		
186	Pro Asp Ile Lys Arg Gly Asn Phe Ala Gln Asp Glu Glu Asp Leu Ile		
187	65 70 75		
189	atc aaa ctc cat gct ctc ctt ggt aac cgg tgg tca ctg ata gct ggt 350		

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/453,387A

DATE: 07/25/2001

TIME: 15:26:46

Input Set : A:\Uco-956.app

Output Set: N:\CRF3\07252001\I453387A.raw

190	Ile	Lys	Leu	His	Ala	Leu	Leu	Gly	Asn	Arg	Trp	Ser	Leu	Ile	Ala	Gly	
191	80						85						90				
193	aga	tta	cca	gga	aga	aca	gat	aat	gaa	gtg	aag	aac	tat	tgg	aat	tcc	398
194	Arg	Leu	Pro	Gly	Arg	Thr	Asp	Asn	Glu	Val	Lys	Asn	Tyr	Trp	Asn	Ser	
195	95						100					105					
197	cat	ata	aag	aga	aag	cta	atg	aag	atg	ggg	atc	gat	cct	aat	aac	cat	446
198	His	Ile	Lys	Arg	Lys	Leu	Met	Lys	Met	Gly	Ile	Asp	Pro	Asn	Asn	His	
199	110						115				120			125			
201	aag	ttg	aac	caa	tat	cct	cat	cat	gtt	ggt	ccc	ctt	aac	ccc	acc	acc	494
202	Lys	Leu	Asn	Gln	Tyr	Pro	His	His	Val	Gly	Pro	Leu	Asn	Pro	Thr	Thr	
203							130				135			140			
205	acc	acc	tcc	atg	gat	gtg	gca	tgt	aag	ctt	aga	gtg	tgt	tca	aca	gac	542
206	Thr	Asn	Ser	Met	Asp	Val	Ala	Cys	Lys	Leu	Arg	Val	Cys	Ser	Thr	Asp	
207	145						150				155						
209	aat	gat	gat	ggg	atc	tca	gat	gct	gca	agt	tat	ctc	gaa	gac	gca	aca	590
210	Asn	Asp	Asp	Gly	Ile	Ser	Asp	Ala	Ala	Ser	Tyr	Leu	Glu	Asp	Ala	Thr	
211	160						165				170						
213	ccg	ccc	act	ggt	ata	tcc	aac	ttg	gac	ctt	gat	ctc	aca	att	gct	ttt	638
214	Pro	Pro	Thr	Gly	Ile	Ser	Asn	Leu	Asp	Leu	Asp	Leu	Thr	Ile	Ala	Phe	
215	175						180				185						
217	cct	tcg	agt	cct	atc	aag	aat	att	att	gaa	gaa	agc	cag	cag	aaa	aca	686
218	Pro	Ser	Ser	Pro	Ile	Lys	Asn	Ile	Ile	Glu	Glu	Ser	Gln	Gln	Lys	Thr	
219	190						195			200			205				
221	gca	tct	att	gta	aca	aat	gat	gaa	gaa	gaa	caa	tat	aca	gtc	cct	acc	734
222	Ala	Ser	Ile	Val	Thr	Asn	Asp	Glu	Glu	Glu	Gln	Tyr	Thr	Val	Pro	Thr	
223							210			215			220				
225	ctt	ctt	ctt	ttc	aga	tga	gacaaaaaaaaaa	aaagcctcac	acatgtggag							782	
226	Leu	Leu	Leu	Phe	Arg												
227				225													
229	attcgtgcaa	aagacctaaa	ggcttacgaa	ggcaacatgc	acgccattgt	caaattcttt											842
231	tggatgatgg	attgaaacca	tatccttgc	cattagaaag	gaggaagata	agctaaaact											902
233	gtattattgt	gtataaaattt	ggtagaaaga	aagatttcaa	cttaagaatt	aggatcaaat											962
235	aactgaatga	atgaacgaat	tgcagataag	ttgttaggag	gttttcaatc	aacttatctg											1022
237	caattaattt	ggtggagctg	atgttaggatg	atgagttcat	cgtacatgaa	ctgaaccttt											1082
239	gatatttcag	gctctaattt	tctgtttgtt	tgcgtaaaga	tattcttcaa	tgtgagatca											1142
241	gctaaaaaaa																1151
244	<210>	SEQ ID NO:	4														
245	<211>	LENGTH:	226														
246	<212>	TYPE:	PRT														
247	<213>	ORGANISM:	Gossypium hirsutum														
249	<400>	SEQUENCE:	4														
250	Met	Arg	Lys	Pro	Cys	Cys	Asp	Lys	Gln	Gly	Thr	Asn	Lys	Gly	Ala	Trp	
251	1				5				10					15			
252	Ser	Lys	Gln	Glu	Asp	Gln	Lys	Leu	Ile	Asp	Tyr	Ile	Arg	Ile	His	Gly	
253					20				25				30				
254	Glu	Gly	Cys	Trp	Arg	Ser	Leu	Pro	Lys	Ala	Ala	Gly	Leu	His	Arg	Cys	
255					35				40				45				
256	Gly	Lys	Ser	Cys	Arg	Leu	Arg	Trp	Ile	Asn	Tyr	Leu	Arg	Pro	Asp	Ile	
257					50				55				60				

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/453,387A

DATE: 07/25/2001

TIME: 15:26:46

Input Set : A:\Uco-956.app

Output Set: N:\CRF3\07252001\I453387A.raw

258 Lys Arg Gly Asn Phe Ala Gln Asp Glu Glu Asp Leu Ile Ile Lys Leu  
 259 65 70 75 80  
 260 His Ala Leu Leu Gly Asn Arg Trp Ser Leu Ile Ala Gly Arg Leu Pro  
 261 85 90 95  
 262 Gly Arg Thr Asp Asn Glu Val Lys Asn Tyr Trp Asn Ser His Ile Lys  
 263 100 105 110  
 264 Arg Lys Leu Met Lys Met Gly Ile Asp Pro Asn Asn His Lys Leu Asn  
 265 115 120 125  
 266 Gln Tyr Pro His His Val Gly Pro Leu Asn Pro Thr Thr Thr Asn Ser  
 267 130 135 140  
 268 Met Asp Val Ala Cys Lys Leu Arg Val Cys Ser Thr Asp Asn Asp Asp  
 269 145 150 155 160  
 270 Gly Ile Ser Asp Ala Ala Ser Tyr Leu Glu Asp Ala Thr Pro Pro Thr  
 271 165 170 175  
 272 Gly Ile Ser Asn Leu Asp Leu Asp Leu Thr Ile Ala Phe Pro Ser Ser  
 273 180 185 190  
 274 Pro Ile Lys Asn Ile Ile Glu Glu Ser Gln Gln Lys Thr Ala Ser Ile  
 275 195 200 205  
 276 Val Thr Asn Asp Glu Glu Gln Tyr Thr Val Pro Thr Leu Leu Leu  
 277 210 215 220  
 278 Phe Arg  
 279 225  
 283 <210> SEQ ID NO: 5  
 284 <211> LENGTH: 1081  
 285 <212> TYPE: DNA  
 286 <213> ORGANISM: Gossypium hirsutum  
 288 <220> FEATURE:  
 289 <221> NAME/KEY: CDS  
 290 <222> LOCATION: (86)..(994)  
 291 <223> OTHER INFORMATION: GhMYB7  
 293 <400> SEQUENCE: 5  
 294 ctccccgcgg tggccgcgc tctagaacta gtggatcccc cgggctgcag gaattcggca 60  
 296 cgaggaaaga agtgtgaaaa aaaaa atg gga agg agt cct tgt tgt tct aag 112  
 297 Met Gly Arg Ser Pro Cys Cys Ser Lys  
 298 1 5  
 300 gaa ggc ctt aac aga gga gct tgg act gct ctt gaa gac aaa att ctt 160  
 301 Glu Gly Leu Asn Arg Gly Ala Trp Thr Ala Leu Glu Asp Lys Ile Leu  
 302 10 15 20 25  
 304 aaa gat tat atc aaa gta cac ggt gaa ggt cgt tgg aga aat ctc ccc 208  
 305 Lys Asp Tyr Ile Lys Val His Gly Glu Gly Arg Trp Arg Asn Leu Pro  
 306 30 35 40  
 308 aaa aga gct ggt ctt aag aga tgt ggg aaa agt tgt agg ctt cgg tgg 256  
 309 Lys Arg Ala Gly Leu Lys Arg Cys Gly Lys Ser Cys Arg Leu Arg Trp  
 310 45 50 55  
 312 ttg aat tat ttg aga cct gat att aaa aga ggt aac ata tca cct gac 304  
 313 Leu Asn Tyr Leu Arg Pro Asp Ile Lys Arg Gly Asn Ile Ser Pro Asp  
 314 60 65 70  
 316 gag gaa gag ctt atc atc aaa ctc cac aaa ctc ttg gga aac aga tgg 352  
 317 Glu Glu Leu Ile Ile Lys Leu His Lys Leu Leu Gly Asn Arg Trp

Use of n and/or Xaa has been detected in the Sequence Listing.  
 → Review the Sequence Listing to insure a corresponding  
 explanation is presented in the <220> to <223> fields of  
 each sequence using n or Xaa.

**VERIFICATION SUMMARY**

PATENT APPLICATION: US/09/453,387A

DATE: 07/25/2001

TIME: 15:26:47

Input Set : A:\Uco-956.app

Output Set: N:\CRF3\07252001\I453387A.raw

L:104 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1  
L:447 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:7  
L:575 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:9  
L:593 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10  
L:798 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:25  
L:827 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:26